PRISM: CONTROLLABLE DIFFUSION FOR COMPOUND IMAGE RESTORATION WITH SCIENTIFIC FIDELITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Scientific and environmental imagery are often degraded by multiple compounding factors related to sensor noise and environmental effects. Existing restoration methods typically treat these compound effects by iteratively removing fixed categories, lacking the compositionality needed to handle real-world mixtures and often introducing cascading artifacts, overcorrection, or signal loss. We present **PRISM** (**P**recision **R**estoration with **I**nterpretable **S**eparation of **M**ixtures), a prompted conditional diffusion framework for *expert-in-the-loop controllable restoration* under compound degradations. PRISM combines (1) compound-aware supervision on mixtures of distortions and (2) a weighted contrastive disentanglement objective that aligns compound distortions with their constituent primitives to enable high-fidelity joint restoration. We outperform image restoration baselines on unseen complex real-world degradations, including underwater visibility, under-display camera effects, and fluid distortions. PRISM also enables selective restoration. Across microscopy, wildlife monitoring, and urban weather datasets, our method enhances downstream analysis by letting experts remove only degradations that hinder analysis, avoiding black-box "over-restoration." Together, these results establish PRISM as a generalizable, controllable framework for high-fidelity restoration in domains where scientific utility is a priority.

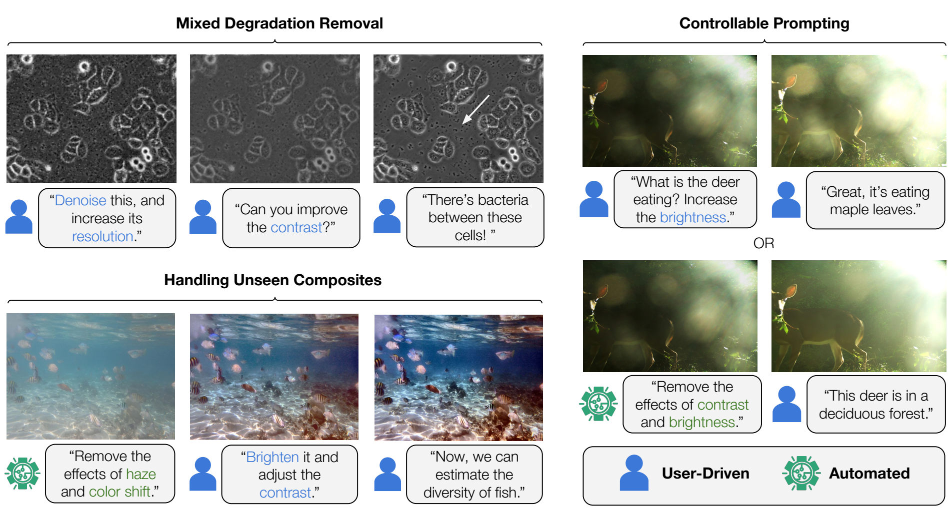

Figure 1: *Expert-in-the-Loop Restoration with PRISM.* PRISM learns separable, compositional embeddings of distortions, enabling robust compound restoration and zero-shot handling of unseen mixtures. It balances high fidelity with expert control, supporting both automatic restoration and prompt-driven, selective correction for scientific analysis.

1 INTRODUCTION

Scientific and environmental imagery is rarely degraded by a single factor. Instead, images typically suffer from *compounding effects* that vary across datasets and collection settings. For instance, underwater images combine low light, scattering, and wavelength-dependent absorption effects (Akkaynak & Treibitz, 2018; Chiang & Chen, 2011), while satellite imagery suffers from overlapping sensor noise, haze, and cloud occlusions (King et al., 2013; Ahmad et al., 2019).

Specialized single-distortion models (e.g., dehazing or cloud removal) enable domain experts to preprocess noisy data before conducting analysis. These approaches are often carefully hand-crafted for specific datasets and distortion types, making them brittle when degradations occur unpredictably. For example, camera trap data may combine the effects of motion blur, weather, and lighting that vary across images from the same deployment, making single-distortion pipelines ineffective. While this has motivated generalist "all-in-one" models (Li et al., 2020; Potlapalli et al., 2023b), evaluations on these methods typically emphasize image quality rather than *scientific fidelity*. In practice, restoration must preserve signals critical for precision and analysis, not just aesthetics.

Moreover, current frameworks fail in two ways: (1) sequential/iterative removal of single distortions, leading to cascading artifacts, or (2) indiscriminate removal, erasing signals that should be preserved. In science, "more restoration" may mean more error: denoising may erase faint galaxies (Starck et al., 2002), and super-resolution can hallucinate microscopy structures (Christensen, 2022). Few models let experts control these tradeoffs.

We argue that restoration for science requires three principles: *simultaneous over sequential correction*, *control over automation*, and *precision over aesthetics*. We introduce PRISM, a conditional diffusion framework that disentangles compound degradations and enables expert-guided, faithful restoration. Our contributions are:

1. A *framework with compound-aware supervision and contrastive disentanglement* across a broad diversity of primitive tasks, producing separable embeddings of distortions for robust sequential and compound restoration, even under unseen real-world mixtures;
2. A *novel benchmark for scientific utility* spanning remote sensing, ecology, biomedical, and urban domains—including our newly-introduced Rooftop Cityscapes dataset—that evaluates task fidelity rather than perceptual scores;
3. A systematic study showing that *controllability is not a convenience but a necessity*: selective restoration significantly improves scientific accuracy under unseen, real-world degradations, establishing precision and expert guidance as core requirements for trustworthy restoration.

2 Related Works

2.1 Restoration in Scientific Domains

Restoration has long been integral to scientific imaging: early astronomy corrected photographic plates (Gull & Daniell, 1978), while biomedical imaging relied on denoising and deblurring for diagnostics (Buades et al., 2005; Dabov et al., 2007). Modern deep learning pipelines continue this pattern with destriping in astronomy surveys (Liu et al., 2025; Vojtekova et al., 2021) or MRI denoising (Yan et al., 2024; Manjón & Coupe, 2018; Kidoh et al., 2020). These methods are effective but assume degradations are fixed and known.

In reality, compound degradations are common. Domain-specific models like Sea-Thru (Akkaynak & Treibitz, 2019) or dark channel priors for atmospheric correction (Li et al., 2018; Guo et al., 2019) explicitly model these effects, but rely on tailored, paired datasets where "ground truth" is often simulated, limiting generalization.

Several works across scientific imaging domains caution that aggressive or over-generalized restoration, particularly where multiple degradations are removed simultaneously, can introduce unwanted artifacts and compromise fidelity. In microscopy, Lu et al. (2025) showed that over-denoising obliterates critical details; in underwater monitoring, Cecilia & Murugan (2022) found generic denoisers over-smooth marine structures, obscuring ecologically relevant edges. These studies highlight that clean images are not always better: over-restoration risks discarding weak but meaningful signals or introducing unwanted artifacts. Taken together, there is a clear need for frameworks that jointly handle compound effects while allowing expert control over what to preserve.

2.2 Compound Degradation Image Restoration

"All-in-one" models share backbones but treat degradations independently, e.g., All-WeatherNet (Li et al., 2020), TransRestorer (Chen et al., 2021), and SmartAssign (Wang et al., 2023). . Universal networks (MT-Restore (Chen et al., 2022b), All-in-OneNet (Li et al., 2022b), PatchDiffuser (Özdenizci & Legenstein, 2023)) improve flexibility, yet fail when degradations overlap nonlinearly. Composite approaches such as OneRestore (Guo et al., 2024) and AllRestorer (Mao et al., 2024) better capture interactions. However, they remain black boxes with no expert-in-the-loop control, and their generalization to unseen mixtures is unclear.

2.3 Prompt-Guided Restoration and Conditional Diffusion

Prompt-based methods provide greater adaptability. The scene descriptor tokens in OneRestore, AllRestorer, and PromptIR (Potlapalli et al., 2023b) demonstrate that textual conditioning improves all-in-one restoration on limited sets of fixed categories. Diffusion-based methods (Ho et al., 2020; Dhariwal & Nichol, 2021; Rombach et al., 2022) extend this idea with higher fidelity: DiffPlugin (Liu et al., 2024) introduces modular prompt conditioning with natural language, MPerceiver (Ai et al., 2024) encodes degradation as "tokens", and AutoDIR (Jiang et al., 2024) prompts over dominant subtasks iteratively. These works either require predefined vocabularies, treat degradations discretely, or accumulate errors in sequential restoration. Jiang et al. (2024) and Ai et al. (2024) hint at primitive decomposition for generalization but stop short of modeling mixtures compositionally.

A recent survey (Jiang et al., 2025) highlights two key gaps: (1) explicitly handling complex, real-world degradations and (2) establishing standardized multi-domain evaluations. We address both by enforcing compositionality in the image embedding space and introducing a Mixed Degradations Benchmark (MDB) and downstream task evaluation pipeline. Furthermore, prior approaches have not jointly addressed (1) controllability, (2) compound degradation removal, and (3) robustness to unseen mixtures. PRISM fills this gap with compound-aware supervision, weighted contrastive disentanglement, and prompt-based conditioning, enabling precise and interpretable restoration.

3 Methods

PRISM is trained on a large-scale diverse dataset of mixed degradations, integrates compound-aware supervision with contrastive disentanglement, and uses a latent diffusion backbone augmented with modules for content preservation and controllable restoration.

3.1 Building a Dataset of Compound Degradations

We construct a synthetic dataset from diverse scientific domains: ImageNet (Deng et al., 2009), Sentinel-2 patches from Sen12MS (Schmitt et al., 2019), iWildCam 2022 (JohnBeuving et al., 2022), EUVP underwater imagery (Islam et al., 2020), CityScapes (Cordts et al., 2016), BioSR microscopy slides (Gong et al., 2021), Brain Tumor MRI (Nickparvar, 2021), and high-resolution Subaru/HSC sky surveys (Miao et al., 2024). Across these datasets, we sample 1.5M clean images that serve as the ground truth targets during training.

Compound-Aware Supervision. Each image is degraded by up to three distortions sampled from a library including geometric warping, blur, photometric shifts and weather effects, etc. We cap distortions at three to capture challenging compound cases while maintaining efficiency and prompt interpretability. For each image I_{clean}, a distorted version is generated as

$$I_{\text{dist}} = d_{i_1} \circ d_{i_2} \circ d_{i_3}(I_{\text{clean}}), \tag{1}$$

where $\circ$ denotes composition. The distortions are applied in random order with varied parameters (i.e. kernel size for blurring, intensity of snowfall, etc.) that determine degradation intensity. Prompts p describing distortions are auto-generated with GPT-4 (Hurst et al., 2024) to simulate the variability in natural-language queries that may be provided as input. We also include *partial* prompts (remove a subset of distortions) and *negative* prompts (remove a non-present distortion), training the model to both handle joint degradations and respect user instructions. The dataset consists of triplets $(I_{\text{clean}}, I_{\text{dist}}, p)$. Further details on distortions, dataset construction, and sampling are provided in Appendix Sec. D.

3.1.1 THE PRISM MODEL

Our framework builds on composite/all-in-one restoration (Guo et al., 2024; Jiang et al., 2024; Ai et al., 2024) but emphasizes *controllability* and *precision* under compound degradations. We first fine-tune the CLIP image encoder, keeping the text encoder frozen, on our mixed-degradation dataset to ensure that embeddings preserve semantic content while becoming distortion-invariant; once adapted, we freeze both CLIP encoders to provide a stable conditioning space for training the latent diffusion backbone.

Disentangling Distortions. Naive CLIP embeddings are poorly suited to restoration, as they cluster images by semantic content rather than image quality. Jiang et al. (2024) showed that quality-aware embeddings improve restoration by shifting focus to degradations. We extend this to the compound distortion setting, encoding compositionality so embeddings reflect both individual and overlapping degradations (e.g., an image with haze+rain sits closer to haze or rain than to noise).

Let I_{clean} be a clean image with m degraded variants, where each variant I^j_{dist} is generated by applying either a single primitive degradation (e.g., haze, blur, noise) or a mixture. With embeddings $f(\cdot)$ and cosine similarity $\mathrm{sim}(\cdot,\cdot)$, each degraded variant I^j_{dist} is aligned with I_{clean} and repelled from (i) sibling degradations I^k_{dist} for $k \neq j$ and (ii) other minibatch images $\mathcal{B}_{\text{other}}$.

To reflect the fact that some degradations are more similar than others, sibling negatives are weighted by the Jaccard distance between their degradation sets, $w_{jk} = \exp\left(1 - \frac{|d^j \cap d^k|}{|d^j \cup d^k|}\right)$.

This weighting means that two variants sharing any primitives (e.g., haze vs. haze+rain) are treated as more similar than variants with little overlap (e.g., haze vs. noise) are pushed further apart. In other words, the loss models compound degradations as mixtures of their primitives, ensuring the embedding space preserves compositionality. The per-variant contrastive loss is

$$\mathcal{L}^{(j)}_{\text{ctr}} = -\log \frac{\exp(\mathrm{sim}(f(I^j_{\text{dist}}), f(I_{\text{clean}}))/\tau)}{\sum_{k \neq j} w_{jk}\, \exp(\mathrm{sim}(f(I^j_{\text{dist}}), f(I^k_{\text{dist}}))/\tau) + \sum_{n \in \mathcal{B}_{\text{other}}} \exp(\mathrm{sim}(f(I^j_{\text{dist}}), f(n))/\tau)},$$

for some temperature τ (see Sec. G in the Appendix for ablation study). This term is averaged over all clean images per batch and their variants. To further bias embeddings toward fidelity, we add a quality-aware regularizer for each clean image

$$\mathcal{L}_{\text{qual}} = \frac{1}{m} \sum_{j=1}^{m} \sum_{c \in d^j} \hat{p}(c \mid I_{\text{clean}}),$$

where d^j is the set of degradations applied to I^j_{dist}, and $\hat{p}(c \mid I_{\text{clean}})$ is the model's predicted probability that the clean embedding exhibits this configuration. When averaged over a batch, this penalizes clean embeddings that encode artifacts from their degraded counterparts.

The final objective is $\mathcal{L}_{\text{CLIP}} = \mathcal{L}_{\text{ctr}} + \mathcal{L}_{\text{qual}}$. This yields embeddings that cluster degraded views with their clean counterpart and maintain compositional structure, enabling precise and controllable restoration.

Prompting. Using this disentangled latent space, PRISM can generate its own input prompts. We predict a distribution over candidate degradations $\hat{\mathcal{D}}$ via multi-label classification on encoder embeddings, converting predictions into prompts of the predefined fixed format. This hybrid design supports both expert-guided and automated restoration. For our experiments below, we use fixed text prompts of the form *"remove the effects of distortions x, y, and z"*.

PRISM bridges natural language prompts to task predictions through a two-stage process. Each prompt from our training set described above is paired with a ground-truth label set from a fixed vocabulary of distortions (blur, haze, blur and haze, etc.), enabling supervised training of the mapping between text and tasks. At inference time, user-provided prompts are encoded with the frozen CLIP text encoder and projected into this label space, producing a multi-label prediction $\hat{\mathcal{D}}$ that specifies

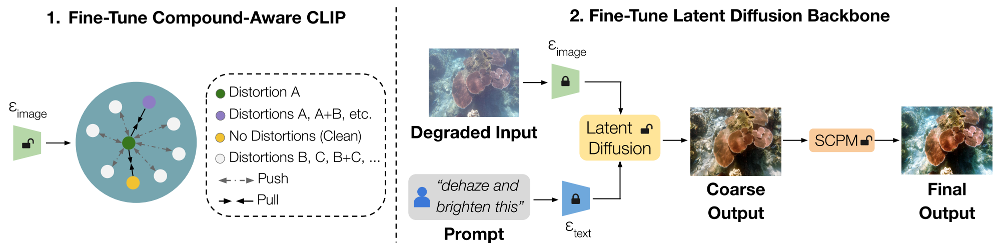

Figure 2: *Overview of PRISM.* We first fine-tune CLIP's image encoder to disentangle image embeddings by distortion. The degraded input and user prompt are then used to condition the latent diffusion backbone, the coarse outputs of which are refined with a Semantic Content Preservation Module (SCPM) to yield the final restored output.

which degradations to remove. This design allows PRISM to accept free-form natural language while ensuring predictions align with a consistent set of restoration tasks.

Model Backbone. We adopt a latent diffusion model (Rombach et al., 2022), operating in VAE-compressed space. Unlike prior sequential restoration methods, which remove distortions step-by-step (Jiang et al., 2024), PRISM conditions diffusion directly on composite prompts, reducing cascading errors. While MPerceiver (Ai et al., 2024) uses concatenated tokens to encode multiple degradations, this method does not explicitly model compositionality or enable controllable prompts.

To recover fine details, we jointly fine-tune a Semantic Content Preservation Module (SCPM) that fuses encoder and decoder features via adaptive modulation

$$f_{\text{refined}} = \gamma(f_{\text{enc}}) \odot \text{Norm}(f_{\text{dec}}) + \beta(f_{\text{enc}}), \tag{2}$$

where $\gamma(\cdot)$ and $\beta(\cdot)$ are learned affine transforms, and $\odot$ denotes element-wise multiplication. Residual and attention blocks process f_{refined} before final decoding by D_{VAE}. By reintroducing encoder features at the decoding stage, SCPM retains fine structures (edges, textures, or small objects) that are often lost in the bottleneck representation.

Refer to Appendix Sec. G for more details and ablations over model components (loss, SCPM, etc.). For fair comparison, all baselines are trained on the same set of primitive distortions. Training details, compute requirements, and baselines are described in Appendix Secs. C and F.

3.2 Evaluation

We evaluate PRISM on: (1) compound and controllable restoration, (2) handling unseen real-world composites, and (3) downstream utility. Unless noted otherwise, we use manual prompting with pre-defined distortion types. Full details on datasets and evaluation are in Appendix D and E.

Mixed Degradations Benchmark (MDB). We build a fixed testbed of triplets $(I_{\text{clean}}, I_{\text{dist}}, p)$ with up to three randomly composed distortions and a matching prompt p. MDB measures sequential vs. composite prompting and prompt faithfulness under compound degradations. This dataset builds off of the CDD-11 proposed by Guo et al. (2024) to span a broader diversity of real-world degradations.

Handling Unseen Distortions. For zero-shot tests, we evaluate on real domains with compound distortions not explicitly seen in training: underwater effects (low light, color, and haze) in UIEB (Li et al., 2019), under-display camera artifacts (low light, blur, and contrast) (Zhou et al., 2021), and fluid-based distortions (refraction and warping) (Thapa et al., 2020). These probe PRISM's ability to extend to novel, physically distinct distortions.

Downstream Utility. Standard benchmarks measure pixel similarity to a clean reference, but this misses whether restored images remain scientifically useful. We instead evaluate restoration through downstream tasks using real datasets with natural distortions and undistorted views as ground truth. To reflect how restoration outputs are typically used in practice, we use *off-the-shelf pretrained models*, giving a conservative but practical measure of utility. We test across four real-world domains:

1. **Remote sensing with Sen12MS (Schmitt et al., 2019):** landcover classification (Papoutsis et al., 2023) on cloudy satellite data, with labels from cloudless samples.

2. **Wildlife monitoring with iWildCam 2022 (JohnBeuving et al., 2022):** species classification with SpeciesNet (Gadot et al., 2024) on low-confidence nighttime images with expert labels from high-confidence frames of the same sequence.
3. **Segmentation and tracking in microscopy with BioSR (Gong et al., 2021):** segmentation of clathrin-coated pits from low signal-to-noise data using MicroSAM (Archit et al., 2025), compared to high quality structured illumination microscopy (SIM) ground truth.
4. **Urban forest monitoring using our novel Rooftop Cityscapes dataset:** panoptic segmentation (Lin et al., 2017) of cityscapes under haze/low light, with paired, labeled clear-weather data. See Appendix Sec. E for details on this custom dataset.

4 Results and Discussion

We present results that evaluate PRISM's ability to restore images degraded by multiple simultaneous distortions, showing how compound-aware supervision and contrastive disentanglement together improve restoration fidelity and controllability under complex degradation scenarios.

4.1 Breaking the Cascade: Compound Restoration Made Robust

Table 1: PRISM outperforms baselines on MDB, where each test image has up to three distortions. Best results are **bolded**, second-best are underlined.

Method	**PSNR ↑**	**SSIM ↑**	**FID ↓**	**LPIPS ↓**
AirNet (Li et al., 2022a)	19.23	0.742	78.55	0.382
Restormer$_A$ (Zamir et al., 2022)	20.84	0.768	70.11	0.365
NAFNet$_A$ (Chen et al., 2022a)	21.51	0.776	68.30	0.352
PromptIR (Potlapalli et al., 2023a)	22.67	0.801	62.78	0.298
OneRestore (Guo et al., 2024)	22.94	0.812	59.42	0.276
DiffPlugin (Liu et al., 2024)	23.45	0.821	53.88	0.255
MPerceiver (Ai et al., 2024)	24.19	0.829	**48.18**	0.235
AutoDIR (Jiang et al., 2024)	23.84	0.833	50.75	0.246
PRISM (ours)	**25.62**	**0.842**	48.97	**0.218**

Sequentially removing distortions often accumulates errors: each step can introduce artifacts, smoothing, or inconsistencies. Restoring all distortions jointly avoids these pitfalls and yields more stable, high-fidelity results. Our MDB evaluation supports this intuition (qualitative results in Appendix Figs. 16 and 15).

Table 1 highlights a divide between early all-in-one models (AirNet, Restormer, NAFNet, PromptIR), which are trained per-distortion and generalize poorly to mixtures, and recent composite/diffusion approaches (DiffPlugin, MPerceiver, AutoDIR). OneRestore struggles with capacity limits, while diffusion methods improve perceptual fidelity but still rely on single-distortion or sequential training.

PRISM achieves the best results across both fidelity (PSNR/SSIM) and perceptual metrics (FID/LPIPS), owing to two design choices: (1) *compound-aware supervision*, which trains on mixed degradations, and (2) *contrastive disentanglement* of embeddings to separate degradations from semantics. We study the contributions of each below.

Compound-aware supervision supports restoration under increasingly complex mixed degradations. Training on combinatorial mixtures of degradations (full, partial, and negative restoration) teaches how they interact, enabling simultaneous removal without cascading errors. PRISM matches baselines on single distortions but excels as complexity grows, even on unseen cases with four distortions. Fig. 3 shows that training on composites explicitly outperforms training on separate distortions, and that improved image embeddings from our contrastive loss provide an additional boost over baselines.

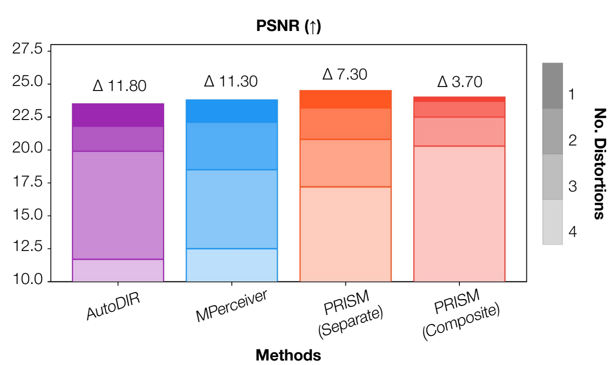

Figure 3: *PRISM trained on composite examples scales best with the number of distortions.* This outperforms our model trained on each degradation separately as well as comparable baselines, emphasized by the Δ (change in performance across test images with 1 vs. 4 distortions) above each bar.

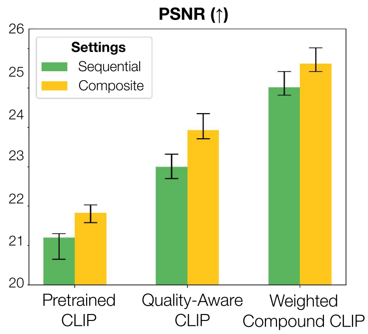

Figure 4: *Latent disentanglement of distortions enables faithful stepwise and single-shot restoration.* This closes the gap between prompting strategies.

Contrastive disentanglement improves both partial and composite restoration. Composite restoration is most effective when distortions remain disentangled in the latent space. Adding only a quality-aware loss improves restoration fidelity, but embeddings still collapse under mixed degradations. By contrast, our compound-aware contrastive loss enforces a compositional geometry, embedding compounds alongside their primitives. This substantially improves composite restoration and narrows the gap between sequential and single-shot prompting (see Fig. 4). Fig. 12 in the Appendix visualizes the effect of the fine-tuning CLIP on the image embedding space, and we provide full versions of Figs. 3 and 4 in Section H.

By combining compound-aware supervision with contrastive disentanglement, PRISM not only outperforms existing baselines on restoring compound degradations, but also enables expert-in-the-loop control. In improving sequential restoration, users can specify which degradations to remove and which signals to preserve. The result is a restoration framework that is both more precise and more flexible, making it well-suited for real-world scientific and environmental imaging applications where overcorrection or artifact introduction can compromise downstream analysis.

4.2 COMPOSITIONALITY ENABLES ADAPTIVE RESTORATION IN NOVEL SETTINGS

Strong compound restoration performance also supports generalization to unseen degradations. If degradations are represented compositionally, then novel composites can be modeled as combinations of known primitives. This means PRISM can automatically identify constituent distortions and restore them, even if the exact combination was never seen in training.

We evaluate zero-shot restoration on three domains: underwater imagery (UIEB), under-display cameras (POLED), and fluid lensing (ThapaSet), all featuring novel complex distortions (see Table 2). We include qualitative results in Appendix Sec. H.

Table 2: PRISM achieves state-of-the-art zero-shot performance across underwater (UIEB), under-display camera (POLED), and fluid lensing (ThapaSet) benchmarks. Best results are **bolded**, second-best are <u>underlined</u>.

Method	UIEB (Li et al., 2019)			POLED (Zhou et al., 2021)			ThapaSet (Thapa et al., 2020)		
	PSNR ↑	**SSIM ↑**	**LPIPS ↓**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
AirNet (Li et al., 2022a)	16.95	0.755	0.312	13.21	0.517	0.705	19.52	0.741	0.386
Restormer$_A$ (Zamir et al., 2022)	17.42	0.771	0.297	14.11	0.533	0.684	21.14	0.782	0.354
NAFNet$_A$ (Chen et al., 2022a)	17.28	0.739	0.309	11.04	0.561	0.719	20.88	0.774	0.363
PromptIR (Potlapalli et al., 2023b)	20.61	0.879	0.183	18.37	0.621	0.512	22.84	0.808	0.282
OneRestore (Guo et al., 2024)	21.41	0.886	0.171	19.28	0.633	0.488	23.76	0.821	0.263
DiffPlugin (Liu et al., 2024)	22.05	0.895	0.162	19.91	0.641	0.469	24.06	0.827	0.252
MPerceiver Ai et al. (2024)	<u>22.74</u>	<u>0.903</u>	<u>0.152</u>	<u>20.36</u>	<u>0.647</u>	0.448	24.39	0.832	<u>0.238</u>
AutoDIR (Jiang et al., 2024)	22.63	<u>0.903</u>	0.153	20.29	0.646	**0.441**	<u>24.52</u>	<u>0.834</u>	0.243
PRISM (ours)	**23.52**	**0.914**	**0.139**	**21.46**	**0.661**	<u>0.444</u>	**25.28**	**0.845**	**0.224**

All-in-one models (AirNet, Restormer, NAFNet) fall short under these conditions, while composite methods like OneRestore are bottlenecked by under-parameterization. Diffusion models (AutoDIR, MPerceiver) generalize better but treat degradations as isolated factors, reducing robustness under unseen compounds.

PRISM's separable, compound-aware latent space drives its state-of-the-art zero-shot performance across diverse domains. Rather than explicitly modeling complex physical processes such as the propagation of ripples over water, PRISM learns a representation where unseen composites align with their constituent primitives (for example, haze+brightness embeddings lie between haze and brightness) preserving compositional structure while remaining distinct from unrelated distortions. By modeling composites both as unique entities and as structured combinations of primitives, the system can interpolate corrections rather than memorizing fixed categories.

Fig. 5 illustrates how this structure translates into practice. In improving performance on partial restoration, PRISM enables experts to issue stepwise prompts (e.g., "unwarp," "fix coloring," "unblur")

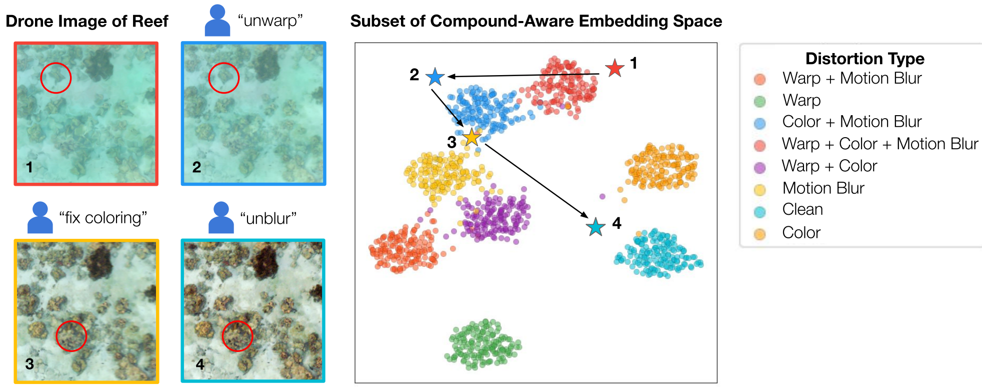

Figure 5: *Controllability supports expert-driven compositional generalization.* Stepwise restoration of a reef drone image shows how prompts progressively target distortions. The embedding visualization demonstrates that unseen compound degradations lie between their constituent primitives, enabling both strong zero-shot performance and expert-in-the-loop control.

to iteratively target different factors for unseen distortion types. This geometry not only enables zero-shot generalization but also provides interpretability and control: experts can explore how distortions relate, refine restoration strategies, and avoid blindly applying black-box corrections. Such interactivity is especially critical in scientific domains, where over-restoration risks introducing artifacts or erasing faint but meaningful signals.

4.3 Prioritizing Precision in Restoration

The results above show that PRISM restores images robustly under compound degradations, both sequentially and in a single step. We next evaluate its impact on *downstream tasks*, shifting from perceptual quality to scientific utility across four domains: remote sensing, ecology, microscopy, and urban monitoring.

Table 3: PRISM faithfully restores data across real scientific datasets. We report mean $\pm$ std on fully-restored outputs over 3 random seeds. Best results are **bolded**, second-best are underlined.

Method	Sentinel-2 (Acc. ↑)	iWildCam (Acc. ↑)	BioSR (mIoU ↑)	Rooftop Cityscapes (mIoU ↑)
AirNet (Li et al., 2022a)	0.791 $\pm$ 0.017	0.961 $\pm$ 0.012	0.523 $\pm$ 0.020	0.635 $\pm$ 0.018
Restormer$_A$ (Zamir et al., 2022)	0.812 $\pm$ 0.015	0.965 $\pm$ 0.011	0.550 $\pm$ 0.017	0.639 $\pm$ 0.016
NAFNet$_A$ (Chen et al., 2022a)	0.807 $\pm$ 0.018	0.970 $\pm$ 0.014	0.546 $\pm$ 0.019	0.635 $\pm$ 0.017
OneRestore (Guo et al., 2024)	0.833 $\pm$ 0.014	0.969 $\pm$ 0.010	0.611 $\pm$ 0.015	0.636 $\pm$ 0.015
PromptIR (Potlapalli et al., 2023b)	0.835 $\pm$ 0.013	0.971 $\pm$ 0.012	0.626 $\pm$ 0.014	0.632 $\pm$ 0.013
DiffPlugin (Liu et al., 2024)	0.838 $\pm$ 0.013	0.975 $\pm$ 0.011	0.643 $\pm$ 0.014	0.636 $\pm$ 0.013
MPerceiver (Ai et al., 2024)	**0.842** $\pm$ **0.012**	0.975 $\pm$ 0.010	0.637 $\pm$ 0.014	0.644 $\pm$ 0.012
AutoDIR (Jiang et al., 2024)	0.841 $\pm$ 0.013	**0.976** $\pm$ **0.012**	0.642 $\pm$ 0.011	0.641 $\pm$ 0.013
PRISM (ours)	**0.842** $\pm$ **0.011**	**0.976** $\pm$ **0.008**	**0.657** $\pm$ **0.012**	**0.650** $\pm$ **0.012**

As shown in Table 3, PRISM consistently yields strong downstream performance, boosting classification in wildlife and remote sensing, and segmentation in microscopy and urban scenes. Gains are especially pronounced for dense segmentation, which depends on fine textures and faint structures often lost under mixed distortions, while classification models are comparatively robust to noise. This highlights PRISM's value for dense prediction tasks where subtle details drive scientific precision.

4.3.1 The Case for Controllability in Science

If a model is powerful enough to remove all degradations, should it always do so? In scientific imaging, the answer is often no. Distortions are frequently entangled with faint but meaningful signals, and indiscriminate restoration can erase these cues or introduce artifacts that mislead downstream analysis. This makes controllability essential: experts must be able to decide which degradations to correct and which to preserve.

As shown in Table 4, *selective controllability significantly improves downstream performance over full restoration* (automatically detecting all distortions present) in three of four domains. In nighttime camera trap data, restoring only contrast improves recognition over full restoration, which can blur subtle texture cues. In microscopy, super-resolution alone aligns best with SIM references, while adding denoising suppresses faint but biologically relevant signals. In urban scenes, removing haze improves segmentation, but also correcting low light over-adjusts vegetation and sky regions. Remote sensing is the exception: full restoration performs slightly better, since removing only clouds leaves images under-illuminated and affected by residual haze.

Table 4: Selective controllability outperforms full restoration across three of four downstream tasks. We report mean ± std over 3 random seeds. Best results are **bolded**.

Domain	Degraded Input	Full Restoration	Selective Restoration	p-value
Remote sensing (Acc. ↑)	0.781 ± 0.013	**0.842 ± 0.011**	0.836 ± 0.012	0.11 (n.s.)
Camera Traps (Acc. ↑)	0.921 ± 0.004	0.976 ± 0.008	**0.984 ± 0.004**	$0.032 < 0.05$
Microscopy (mIoU ↑)	0.353 ± 0.015	0.475 ± 0.012	**0.580 ± 0.010**	$0.018 < 0.05$
Urban scenes (mIoU ↑)	0.548 ± 0.018	0.615 ± 0.014	**0.650 ± 0.012**	$0.041 < 0.05$

Fig. 6 illustrates a use case of controllability in microscopy: super-resolution alone improves segmentation alignment with SIM ground truth, but additional denoising erases faint, biologically relevant structures. Additional examples across other domains are included in Appendix Sec. H.

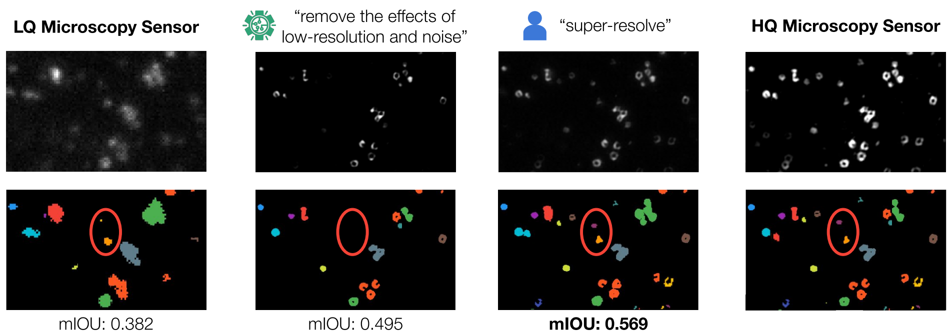

Figure 6: *Selective restoration improves segmentation of clathrin-coated pits in microscopy.* Super-resolution alone improves mIoU, while automatically detecting and removing noise suppresses faint but biologically relevant signals (see regions encircled in red), reducing accuracy.

These results suggest that in scientific domains, controlled restoration allows experts to prioritize downstream utility. In some cases, running inference on the original noisy data is preferable to relying on incorrectly "cleaned" outputs, particularly when downstream models are already robust to noise. Effective restoration systems must therefore prioritize fidelity and expert guidance to ensure outputs remain scientifically reliable.

At the same time, important challenges remain for PRISM. Our training still depends on synthetic augmentations that cannot fully capture real distortions. Moreover, extending controllability beyond "which distortions to remove" toward specifying their intensity and spatial extent would enable localized restoration and finer-grained preservation of scientific signals.

5 Conclusions

Our results show that controllable, compound-aware restoration is critical for scientific and environmental imaging. PRISM outperforms both specialized and generalist baselines by combining (1) compound-aware supervision, which exposes the model to overlapping degradations, and (2) weighted contrastive disentanglement, which organizes the latent space so composite distortions align with their constituent primitives. Together, these yield more robust and interpretable restoration.

We also find strong generalization beyond curated training sets. PRISM achieves robust zero-shot performance on underwater imaging, under-display camera correction, and fluid lensing, showing that compositional representations extend naturally to unseen domains. Importantly, evaluations on real-world composite degradations confirm generalization beyond our synthetic training pipeline.

A key insight is that *more restoration is not always better*. Across diverse domains, we show that indiscriminate removal of degradations suppresses faint but meaningful signals or introduced artifacts. Allowing experts to choose which degradations to correct is essential for scientific precision.

ETHICS STATEMENT

This work builds on publicly available datasets and synthetic distortions, and does not involve human subjects, personally identifiable information, or sensitive data. Where real-world ecological and scientific datasets are used (e.g., remote sensing, microscopy, ecological monitoring), we follow the original licenses, usage guidelines, and citation practices specified by dataset creators.

Our method introduces a controllable restoration framework designed primarily for scientific imaging domains such as ecological monitoring, microscopy, and remote sensing. By enabling expert-guided restoration, our approach can enhance the fidelity and interpretability of critical datasets used to study biodiversity, climate change, and human health. We believe that this has positive societal impact by empowering researchers and practitioners with more reliable tools for environmental stewardship, medical discovery, and other areas where accurate imaging is essential.

At the same time, we acknowledge potential risks. Restoration models can, if misapplied, introduce artifacts that distort scientific findings or be misused in settings such as surveillance, evidence tampering, or misinformation. To mitigate this, we design our system around transparency and user control, highlight the risks of over-restoration, and release evaluation protocols that quantify fidelity under compound degradations. Our framework is not intended for use in high-stakes decision-making without domain expert oversight.

We further note that access to high-quality restoration tools can help democratize science, particularly in resource-constrained regions where imaging equipment may be limited. By making our code and evaluation data publicly available, we aim to broaden participation in ecological and biomedical research. We are committed to responsible dissemination, and to adhering to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

Our Anonymous Github repository contains source code for this work, including end-to-end pipelines for data generation (with metadata for reproducibility), model training, and inference and evaluation across standard and downstream testbeds: `https://anonymous.4open.science/r/PRISM-E4E3/README.md`. Detailed descriptions of our model architecture, training objectives, and evaluation pipelines are provided in Secs. 3 and 4 of the main paper. Hyperparameters, dataset splits, and preprocessing steps are reported in Appendix Secs. C and D. For all baselines, we use publicly available implementations and follow metadata provided by the original papers. Evaluation pipelines and metrics are described in detail in the Appendix Sec. E. We also present extensive ablations and full results in the Appendix Sec. G, justifying each design choice so that other researchers can replicate our findings. Finally, we release our benchmarking datasets for mixed degradation removal and downstream analysis. Together, these resources provide a complete framework for reproducing our results.

REFERENCES

Asmala Ahmad, Shaun Quegan, S Firdaus, Hamzah Sakidin, and M Mawardy. Haze effects on satellite remote sensing imagery and their corrections. *International Journal of Advanced Computer Science and Applications*, 10(10):69–76, 2019.

Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25432–25444, 2024.

Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6723–6732, 2018.

Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1682–1691, 2019.

Anwai Archit, Luca Freckmann, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Carolin Teuber, Melanie Spitzner, Constanza Tapia Contreras, et al. Segment anything for microscopy. *Nature Methods*, 22(3):579–591, 2025.

Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Image denoising by non-local averaging. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 2, pp. ii–25. IEEE, 2005.

Mary Cecilia and S Sakthivel Murugan. Denoising, edge aware restoration and enhancement of single shallow coastal water image. *Fluctuation and Noise Letters*, 21(01):2250009, 2022.

Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12299–12310, 2021.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pp. 17–33, 2022a.

Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, pp. 17653–17662, 2022b.

John Y Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4):1756–1769, 2011.

Charles Nicklas Christensen. *Deep Learning for Image Processing in Optical Super-Resolution Microscopy*. PhD thesis, University of Cambridge (United Kingdom), 2022.

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16 (8):2080–2095, 2007.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5866–5878, 2020.

Tomer Gadot, Ștefan Istrate, Hyungwon Kim, Dan Morris, Sara Beery, Tanya Birch, and Jorge Ahumada. To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Computer Vision*, 18(8):1193–1208, 2024.

Dawei Gong, Tengfei Ma, Julian Evans, and Sailing He. Deep neural networks for image super-resolution in optical microscopy by using modified hybrid task cascade u-net. *Prog. Electromagn. Res*, 171:185–199, 2021.

Stephen F Gull and Geoff J Daniell. Image reconstruction from incomplete and noisy data. *Nature*, 272(5655):686–690, 1978.

Qiang Guo, Hai-Miao Hu, and Bo Li. Haze and thin cloud removal using elliptical boundary prior for remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9124–9137, 2019.

Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *European Conference on Computer Vision*, pp. 255–272. Springer, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020.

Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *European Conference on Computer Vision*, pp. 340–359. Springer, 2024.

JohnBeuving, Phil Culliton, sbeery, and Ștefan Istrate. iwildcam 2022 - fgvc9. `https://kaggle.com/competitions/iwildcam2022-fgvc9`, 2022.

Masafumi Kidoh, Kensuke Shinoda, Mika Kitajima, Kenzo Isogawa, Masahito Nambu, Hiroyuki Uetani, Kosuke Morita, Takeshi Nakaura, Machiko Tateishi, Yuichi Yamashita, et al. Deep learning based noise reduction for brain mr imaging: tests on phantoms and healthy volunteers. *Magnetic resonance in medical sciences*, 19(3):195–206, 2020.

Michael D King, Steven Platnick, W Paul Menzel, Steven A Ackerman, and Paul A Hubanks. Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE transactions on geoscience and remote sensing*, 51(7):3826–3852, 2013.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pp. 17452–17462, 2022a.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17452–17462, 2022b.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *TIP*, 29:4376–4389, 2019.

Jiayuan Li, Qingwu Hu, and Mingyao Ai. Haze and thin cloud removal via sphere model improved dark channel prior. *IEEE Geoscience and Remote Sensing Letters*, 16(3):472–476, 2018.

Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3175–3185, 2020.

Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1925–1934, 2017.

Tie Liu, Yuhui Quan, Yingna Su, Yang Guo, Shu Liu, Haisheng Ji, Qi Hao, Yulong Gao, Yuxia Liu, Yikang Wang, et al. Astronomical image denoising by self-supervised deep learning and restoration processes. *Nature Astronomy*, pp. 1–8, 2025.

Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4197–4208, 2024.

Yuer Lu, Yongfa Ying, Chen Lin, Yan Wang, Jun Jin, Xiaoming Jiang, Jianwei Shuai, Xiang Li, and Jinjin Zhong. Unet-att: a self-supervised denoising and recovery model for two-photon microscopic image. *Complex & Intelligent Systems*, 11(1):55, 2025.

José V Manjón and Pierrick Coupe. Mri denoising using deep learning. In *International workshop on patch-based techniques in medical imaging*, pp. 12–19. Springer, 2018.

Jiawei Mao, Yu Yang, Xuesong Yin, Ling Shao, and Hao Tang. Allrestorer: All-in-one transformer for image restoration under composite degradations. *arXiv preprint arXiv:2411.10708*, 2024.

Jiawei Miao, Liangping Tu, Bin Jiang, Xiangru Li, and Bo Qiu. Astrosr: A data set of galaxy images for astronomical superresolution research. *The Astrophysical Journal Supplement Series*, 274(1):7, 2024.

Msoud Nickparvar. Brain tumor mri dataset, 2021. URL `https://www.kaggle.com/dsv/2645886`.

Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE TPAMI*, 45(8):10346–10357, 2023.

Ioannis Papoutsis, Nikolaos Ioannis Bountos, Angelos Zavras, Dimitrios Michail, and Christos Tryfonopoulos. Benchmarking and scaling of deep learning models for land cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195:250–268, 2023.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090*, 2023a.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023b.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu. Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *arXiv preprint arXiv:1906.07789*, 2019.

Jean-Luc Starck, Eric Pantin, and Fionn Murtagh. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051, 2002.

Simron Thapa, Nianyi Li, and Jinwei Ye. Dynamic fluid surface reconstruction using deep neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21–30, 2020.

Antonia Vojtekova, Maggie Lieu, Ivan Valtchanov, Bruno Altieri, Lyndsay Old, Qifeng Chen, and Filip Hroch. Learning to denoise astronomical images with u-nets. *Monthly Notices of the Royal Astronomical Society*, 503(3):3204–3215, 2021.

Yinglong Wang, Chao Ma, and Jianzhuang Liu. Smartassign: Learning a smart knowledge assignment strategy for deraining and desnowing. In *CVPR*, pp. 3677–3686, 2023.

Xu Yan, MingXuan Xiao, Weimin Wang, Yufeng Li, and Fei Zhang. A self-guided deep learning technique for mri image noise reduction. *Journal of Theory and Practice of Engineering Science*, 4(01):109–117, 2024.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pp. 5728–5739, 2022.

Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *CVPR*, pp. 9179–9188, 2021.